REVISITING (k,ℓ)-STEP METHODS*

MARTIN H. GUTKNECHT†

Dedicated to Lothar Reichel on the occasion of his 60th birthday

Abstract. In the mid-1980s B.N. Parsons [29] and the author [10] independently had the idea to generalize linear stationary k-step methods to stationary (k,ℓ)-step methods, which were further generalized to nonstationary and even nonlinear (k,ℓ)-step methods. Later, conjugate-gradient-type methods that are (k,ℓ)-step methods of a similar sort were introduced and investigated in the PhD thesis of Teri Barth [3]. Recently, the family of Induced Dimension Reduction (IDR) methods [32] aroused some interest for the class of linear nonstationary (k,ℓ)-step methods because IDR(s) fits into it and belongs to a somewhat special subclass [14]. In this paper we first reformulate and review the class of linear nonstationary (k,ℓ)-step methods and a basic theoretical result obtained in [10]. Then we introduce alternative iterations that can be used to implement them and compare them with the iterations suggested and investigated by Barth.

Key words. Iterative method, linear equations, nonlinear equations, Krylov subspace method, (k,ℓ)–step method, semiiterative method

1. Introduction. In [10], the class of (k,ℓ)–step methods was introduced for the abstract framework of solving a fixed point equation $\mathbf{x} = \Phi\,\mathbf{x}$ of a Fréchet differentiable self-mapping Φ defined on some subset $\mathcal{D}_\Phi$ of a complex Banach space $(\mathcal{B}, \|.\|)$. The main target, however, were the cases where $\mathcal{B} := \mathbb{C}^N$ (or even $\mathcal{B} := \mathbb{R}^N$) and Φ is an affine mapping $\Phi\,\mathbf{x} := \mathbf{F}\,\mathbf{x} + \mathbf{b}$ with a square matrix $\mathbf{F} \in \mathbb{C}^{N\times N}$ whose spectrum does not contain 1. Alternatively, one may choose a formulation with a nonlinear equation $\Psi\mathbf{x} = \mathbf{o}$ and a nonsingular linear system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$:

$$\mathbf{x} = \Phi\,\mathbf{x} \qquad \Longleftrightarrow \qquad \Psi\mathbf{x} = \mathbf{o} \qquad \text{if} \quad \Psi :\equiv \mathbf{I} - \Phi\,, \tag{1.1}$$

$$\mathbf{x} = \mathbf{F} + \mathbf{b}\,\mathbf{x} \qquad \Longleftrightarrow \qquad \mathbf{A}\mathbf{x} = \mathbf{b} \qquad \text{if} \quad \mathbf{A} :\equiv \mathbf{I} - \mathbf{F}\,. \tag{1.2}$$

Interestingly, already in the 1960 book of Faddeev and Faddeeva [5] the parallel treatment of iterative methods applied to both the linear fixed point equation and the standard form of a linear equation in $\mathbb{R}^N$ was stressed. In [10] and [29] both paradigm were mentioned, but the derivation of the methods and the theoretical results were based on the fixed point equation paradigm. The latter is also true for the foregoing work of Niethammer, Schemp, and Varga [24, 26, 27], Gutknecht, Niethammer, and Varga [16], and Gutknecht and Kaiser [15] on k–step methods, which long before had been labeled as "semi-iterative" [34]. Despite the well-known fact that these methods created approximations $\mathbf{x}_n$ from an affine Krylov space, that is,

$$\mathbf{x}_n - \mathbf{x}_0 \in \mathcal{K}_n\,, \quad \text{where} \quad \mathcal{K}_n :\equiv \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \mathsf{span}\,(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0)\,,$$

some of the most helpful tools in the treatment of such methods were not optimally used in the above mentioned publications except [10]. We mention in particular the introduction of the matrices that contain the iterates, the residuals, or the search directions as their columns and which allow us to write the total of all iterations in compact notation. An example is the so-called Arnoldi relation which gathers all iterations of the Arnoldi process and yields the interpretation of the Hessenberg matrix that contains the recurrence coefficients as a projection of $\mathbf{A}$; see, e.g., [30].

---

*__Version of May 7, 2014.__ A preliminary version of Sections 1–4 of this publication has been distributed to the participants of the 2nd International Kyoto Forum on Krylov Subspace Methods, March 2010; but it has not been made available to non-participants.

†Seminar for Applied Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland (mhg@math.ethz.ch).

This relation is, e.g., also the key to a simple derivation of the MinRes and GMRes methods [28, 31]. In a more general situation, it can be traced back to Hessenberg [19], so it is appropriate to refer to such a relationship as a Hessenberg relation in general. Analogous "shorthand notation" was also the key concept in describing the relations between various BiCG algorithms [13] and various GCG and GCR algorithms [12].

In [10], such "shorthand notation" was used to describe the recurrences for the iterates and the relationships between equivalent recurrences (defined below in § 3). But, firstly, in some details, the treatment differed from how we would do it now. Secondly, there was no emphasis on the residual recurrences, which are fundamental for the analysis of the methods, though they are not really needed in the implementation of parameter-dependent (k,ℓ)–step methods. (To some extent this was compensated by looking at the recurrences for the residual polynomials.) Thirdly, the presentation in [10] was mainly based on the fixed point equation paradigm, while, at least in the linear case, we definitely prefer the $\mathbf{Ax} = \mathbf{b}$ form now. So, our goal here is to present these methods in the notation and with the tools that we naturally would apply today, nearly 30 years later.[1] Furthermore, using and generalizing a terminology introduced by Young and Jea [35] we notice that the resulting recurrences are of OrthoRes-type, but we can readily introduce normally equivalent recurrences that are of OrthoMin-type or OrthoDir-type. It turns out that short recurrences are preserved by the transformation to OrthoMin-type, but, surprisingly, not for OrthoDir-type. On the other hand, Barth and Manteuffel [3, 2] investigated thoroughly similar short OrthoDir-type recurrences for CG-type methods (satisfying a possible formal residual or error minimizing property), which they called (s,t)–recursions. So, we conclude that in general their (s,t)–recursions are not equivalent to short (k,ℓ)–step methods, and vice versa. This is of particular interest because the short recurrences for unitary matrices introduced 1982 by Gragg [7, 8] and adapted to shifted unitary matrices by Jagels and Reichel [21, 22] can be seen to be (s,t)–recursions with $(s,t) = (0,1)$ and $(s,t) = (1,1)$, respectively [3, 2].

2. Definition of (k,ℓ)–step methods. In the framework of a possibly nonlinear fixed point equation $\mathbf{x} = \Phi\mathbf{x}$ a (k,ℓ)–step method is defined by a recurrence of the form

$$\mathbf{x}_m := \left[\Phi\left(\sum_{i=1}^{\min\{\ell,m\}} \mathbf{x}_{m-i}\gamma_{m-i,m-1}\right) - \sum_{j=1}^{\min\{k,m\}} \mathbf{x}_{m-j}\beta_{m-j,m}\right]\frac{1}{\beta_{m,m}}, \qquad (2.1)$$
$$m = 1, 2, \ldots, M(\leq \infty),$$

with

$$\beta_{m,m} \neq 0, \qquad \sum_{j=0}^{\min\{k,m\}} \beta_{m-j,m} = 1, \quad m = 1, \ldots, M, \qquad (2.2a)$$

$$\gamma_{m-1,m-1} \neq 0, \qquad \sum_{i=1}^{\min\{\ell,m\}} \gamma_{m-i,m-1} = 1, \quad m = 1, \ldots, M. \qquad (2.2b)$$

Here k, ℓ, and M are positive integers, except that we allow them to be ∞. For example, an (∞,∞)–step method is one where in both sums of (2.1) the upper index

[1] The author started the project in Fall 1985 and had a first, much shorter version of the paper (culminating in the equivalence theorem discussed in § 3 below) ready in March 1985, at the end of his eight-month visit at the IBM Thomas Watson Research Laboratory in Yorktown Heights, NY, USA; but then the further work got delayed, so [10] was submitted in May 1988 only.

bound is m, which means that in both sums all previously computed iterates $\mathbf{x}_{m-i}$ back to $\mathbf{x}_0$ may appear. Similarly, an $(\infty,1)$–step method is one where in the first sum only the term with $\mathbf{x}_{m-1}$ exists, while in the second sum all previously computed iterates may appear. The conditions (2.2a) and (2.2b) guarantee that any fixed point of the equation is also a fixed point of the iteration.

If k and ℓ are finite, the number of steps M will normally exceed them, so we will assume that $M \geq \max\{k,\ell\}$. Formally, every (k,ℓ)–step method that converges in at most M steps can be appended by simple steps $\mathbf{x}_m := \Phi(\mathbf{x}_{m-1})$ for $m > M$ if $M \neq \infty$ in the original definition. So, we may assume that $M = \infty$ when convenient, although the iteration may converge in a finite number of steps.

Every (k,ℓ)–step method requires just one evaluation of the function Φ in each step. On the other hand, the amount of memory required depends essentially only on $\max\{k,l\}$. So, the challenge is to find fast converging methods where $\max\{k,l\}$ is small, that is, methods with short recurrences. Once we will have translated recurrence (2.1) into a form suitable for $\mathbf{Ax} = \mathbf{b}$, we will see that nearly all well-known Krylov subspace methods for linear systems have $\ell = 1$, while k is either 1, 2, or ∞. In particular, CG and BiCG are $(2,1)$–step methods.

3. "Shorthand notation" and the Equivalence Theorem. The coefficients $\beta_{m-j,m}$ and $\gamma_{m-i,m-1}$ that appear in (2.1), supplemented by $\beta_{0,0} :\equiv 1$, can be gathered into banded upper triangular square matrices (filled up with zeros) of the order $M+1$:

$$\mathbf{B} :\equiv (\beta_{j,m})_{j,m=0}^{M}, \qquad \mathbf{C} :\equiv (\gamma_{i,m})_{i,m=0}^{M}.$$

Actually, to $\mathbf{C}$ we added an extra column that does not yet appear in (2.1), but will be needed in the sequel.

Note that $\beta_{m-j,m} = 0$ for $j > k$ and $\gamma_{m-i,m} = 0$ for $i \geq \ell$, so $\mathbf{B}$ has upper bandwidth k while $\mathbf{C}$ has upper bandwidth $\ell - 1$ (total bandwidth $k+1$ and ℓ, respectively). From now on we need no longer express the finiteness of k or ℓ explicitly in all formulas. Due to $\beta_{m,m} \neq 0$ and $\gamma_{m,m} \neq 0$ both matrices are nonsingular, and due to their triangularity, any leading principal submatrix of their inverses can be computed from the leading principal submatrix of the same size of $\mathbf{B}$ and $\mathbf{C}$, respectively, even if $M = \infty$. Finally, according to (2.2a) and (2.2b), $\mathbf{B}$ and $\mathbf{C}$ have column sums 1, a property one can express by $\mathbf{e}^{\mathsf{T}}\mathbf{B} = \mathbf{e}^{\mathsf{T}}$, $\mathbf{e}^{\mathsf{T}}\mathbf{C} = \mathbf{e}^{\mathsf{T}}$, where $\mathbf{e} :\equiv \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^{\mathsf{T}}$. In [10] we called finite or infinite nonsingular upper triangular matrices with this property sequence transformation (ST) matrices. It is easy to verify that those of fixed order form a group under multiplication.

As in [10] we further introduce the auxiliary iterates

$$\mathbf{y}_0 :\equiv \mathbf{x}_0\,, \qquad \mathbf{y}_m :\equiv \sum_{i=0}^{m} \mathbf{x}_i \gamma_{i,m}\,, \quad m = 1,\dots,M, \tag{3.1}$$

$$\mathbf{z}_0 :\equiv \mathbf{x}_0\,, \qquad \mathbf{z}_m :\equiv \Phi\mathbf{y}_{m-1}\,, \quad m = 1,\dots,M, \tag{3.2}$$

and the "formal matrices"

$$\begin{aligned}
\mathbf{X} &:\equiv \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \dots & \mathbf{x}_M \end{bmatrix}, \\
\mathbf{Y} &:\equiv \begin{bmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \dots & \mathbf{y}_M \end{bmatrix}, \\
\mathbf{Z} &:\equiv \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_1 & \dots & \mathbf{z}_M \end{bmatrix}, \\
\Phi(\mathbf{Y}) &:\equiv \begin{bmatrix} \mathbf{y}_0 & \Phi\mathbf{y}_0 & \Phi\mathbf{y}_1 & \dots & \Phi\mathbf{y}_M \end{bmatrix},
\end{aligned}$$

where on the last line we have at left a boldface upper case phi, which does not just mean a componentwise application of Φ, but also incorporates a shift in the index.[2] In the general setting, the columns of these "formal matrices" are elements of $\mathcal{B}$, but when $\mathcal{B} = \mathbb{C}^N$ the "formal matrices" become ordinary complex $N \times (M+1)$ matrices. Moreover, in the affine case where $\Phi\,\mathbf{y} = \mathbf{F}\,\mathbf{y} + \mathbf{b}$, we have

$$\mathbf{\Phi}(\mathbf{Y}) = \mathbf{F}\,\mathbf{Y}\,\mathbf{S} + \mathbf{E_b}\,, \tag{3.3}$$

where $\mathbf{E_b} :\equiv \begin{bmatrix} \mathbf{y}_0 & \mathbf{b} & \mathbf{b} & \dots & \mathbf{b} \end{bmatrix}$, and where

$$\mathbf{S} :\equiv \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ 0 & 0 & 0 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

denotes a right-shift matrix of appropriate size; here of size $(M+1)\times(M+1)$. Finally, we denote by $\mathbf{D}$ the diagonal of $\mathbf{B}$ and by

$$\mathbf{G} :\equiv \mathbf{B} - \mathbf{D}$$

the strictly upper triangular part of $\mathbf{B}$.

With all this notation we quickly see that recursion (2.1), which defines a (k,ℓ)–step method, implies the three identities

$$\mathbf{Y} = \mathbf{X}\mathbf{C}\,, \qquad \mathbf{Z} = \mathbf{\Phi}(\mathbf{Y})\,, \qquad \mathbf{X}\mathbf{B} = \mathbf{Z}\,, \tag{3.4}$$

the last of which can be written as

$$\mathbf{X}\mathbf{D} = \mathbf{Z} - \mathbf{X}\mathbf{G}\,. \tag{3.5}$$

Eliminating $\mathbf{Y}$ and $\mathbf{Z}$ from these three identities yields a "shorthand form" of the recursion (2.1) for $\mathbf{x}_m$,

$$\mathbf{X} = [\mathbf{\Phi}(\mathbf{X}\mathbf{C}) - \mathbf{X}\mathbf{G}]\,\mathbf{D}^{-1}\,. \tag{3.6}$$

Likewise, eliminating $\mathbf{X}$ and $\mathbf{Y}$ yields the "shorthand form" of a recursion for $\mathbf{z}_m$,

$$\mathbf{Z} = \mathbf{\Phi}(\mathbf{Z}\widehat{\mathbf{C}})\,, \qquad \text{where} \quad \widehat{\mathbf{C}} :\equiv \mathbf{B}^{-1}\mathbf{C}\,, \tag{3.7}$$

while eliminating $\mathbf{X}$ and $\mathbf{Z}$ leads first to $\mathbf{Y}\mathbf{C}^{-1}\mathbf{B} = \mathbf{\Phi}(\mathbf{Y})$. Then splitting up $\widehat{\mathbf{B}} :\equiv \mathbf{C}^{-1}\mathbf{B}$ into its diagonal part $\widehat{\mathbf{D}}$ and its strictly upper triangular part $\widehat{\mathbf{G}}$, we obtain the "shorthand form" of a recursion for $\mathbf{y}_m$,

$$\mathbf{Y} = \left[\mathbf{\Phi}(\mathbf{Y}) - \mathbf{Y}\widehat{\mathbf{G}}\right]\widehat{\mathbf{D}}^{-1}\,, \qquad \text{where} \quad \widehat{\mathbf{D}} + \widehat{\mathbf{G}} :\equiv \widehat{\mathbf{B}} :\equiv \mathbf{C}^{-1}\mathbf{B}\,. \tag{3.8}$$

Note that due to the group property $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{C}}$ are also ST-matrices.

Clearly, (3.7) and (3.8) are special cases of (3.6), namely with $\mathbf{B} = \mathbf{I}$, $\mathbf{C} = \widehat{\mathbf{C}}$ (i.e., $\mathbf{D} = \mathbf{I}$, $\mathbf{G} = \mathbf{O}$) in case of (3.7) and with $\mathbf{B} = \widehat{\mathbf{B}}, \mathbf{C} = \mathbf{I}$ (i.e., $\mathbf{D} = \widehat{\mathbf{D}}$, $\mathbf{G} = \widehat{\mathbf{G}}$) in case of (3.8) .

[2] In [10], where we did not use boldface fonts for the elements of $\mathcal{B}$ and for the coefficient matrices $\mathbf{B}$ and $\mathbf{C}$, these four "formal matrices" were denoted by $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{\Phi}(\mathbf{y})$.

Expressing the recurrences for $\mathbf{z}_m$ and $\mathbf{y}_m$ in the usual way, we obtain

$$\mathbf{z}_m := \Phi\left(\sum_{i=0}^{m-1} \mathbf{z}_i \widehat{\gamma}_{i,m-1}\right), \qquad m = 1, 2, \ldots, M, \tag{3.9}$$

and

$$\mathbf{y}_m := \left[\Phi(\mathbf{y}_{m-1}) - \sum_{j=0}^{m-1} \mathbf{y}_j \widehat{\beta}_{j,m}\right] \frac{1}{\widehat{\beta}_{m,m}}, \qquad m = 1, 2, \ldots, M. \tag{3.10}$$

Here, $\widehat{\beta}_{j,m}$ and $\widehat{\gamma}_{j,m}$ are the elements of $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{C}}$, respectively.

Definition [10]. Two (∞,∞)–step methods (1.4) are called equivalent if, for any suitable mapping Φ and any suitable starting point $\mathbf{x}_0$, the sequence

$$\mathbf{y}_m := \sum_{i=0}^{m} \mathbf{x}_i \gamma_{i,m}, \qquad m = 0, 1, \ldots, M,$$

that Φ is applied to is identical for both methods if both are started with the same $\mathbf{x}_0$.

Of course, the sequences $\mathbf{z}_m = \Phi(\mathbf{y}_m)$ produced by two equivalent methods will also be identical, but as pointed out in [10], this need not be true for the sequences $\mathbf{x}_m$ that they produce. Surprisingly, even if the iterations converge, $\mathbf{x}_m$ may converge to another limit than $\mathbf{y}_m$ and $\mathbf{z}_m$. For example, if $\mathbf{B} = \mathbf{C}$, then $\mathbf{y}_m = \mathbf{z}_m$, and $\mathbf{y}_m$ is obtained by Picard iteration $\mathbf{y}_m = \Phi(\mathbf{y}_{m-1})$; but if we choose $\mathbf{B} = \mathbf{C}$ with diagonal elements $1/m$, first row elements $(m-1)/m$, and zeros elsewhere, then $\mathbf{x}_m \to \mathbf{x}_0$ independent of the convergence or divergence of the Picard iteration.

One of the main results of [10] was the following Equivalence Theorem, which results now easily from the foregoing:

THEOREM 1. *To every (∞,∞)–step method as defined by (2.1) (or, in short, by (3.6)), where $\mathbf{B} = (\beta_{j,m})_{j=0}^{M}$ and $\mathbf{C} = (\gamma_{i,m})_{i=0}^{M}$ are ST-matrices, there exists an equivalent $(1,\infty)$–step method of the form (3.9), which can be expressed by (3.7), where $\widehat{\mathbf{C}} = \mathbf{B}^{-1}\mathbf{C}$ is again an ST-matrix, and there also exists an equivalent $(\infty,1)$–step method of the form (3.10), which can be expressed by (3.8), where $\widehat{\mathbf{B}} = \widehat{\mathbf{D}} + \widehat{\mathbf{G}} = \mathbf{C}^{-1}\mathbf{B}$ is also an ST-matrix.*

Conversely, if $\widehat{\mathbf{C}}$ is an ST-matrix defining an $(1,\infty)$–step method, and if

$$\widehat{\mathbf{C}} = \mathbf{B}^{-1}\mathbf{C}$$

is any factorization of $\widehat{\mathbf{C}}$ into the inverse of an ST-matrix $\mathbf{B}$ times an ST-matrix $\mathbf{C}$, then $\mathbf{B}$ and $\mathbf{C}$ define an equivalent (∞,∞)–step method (2.1). Likewise, if $\widehat{\mathbf{B}}$ is an ST-matrix defining an $(\infty,1)$–step method, and if

$$\widehat{\mathbf{B}} = \mathbf{C}^{-1}\mathbf{B}$$

is any factorization of $\widehat{\mathbf{B}}$ into the inverse of an ST-matrix $\mathbf{C}$ times an ST-matrix $\mathbf{B}$, then $\mathbf{B}$ and $\mathbf{C}$ define again an equivalent (∞,∞)–step method (2.1).

In particular, to every $(1,\infty)$–step method there belongs an equivalent $(\infty,1)$–step method, the two being related by $\widehat{\mathbf{B}} = \widehat{\mathbf{C}}^{-1}$.

Two (∞,∞)–step methods given by $\mathbf{B}$, $\mathbf{C}$ and $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{C}}$, respectively, are equivalent if and only if $\mathbf{C}^{-1}\mathbf{B} = \widetilde{\mathbf{C}}^{-1}\widetilde{\mathbf{B}}$, i.e., if and only if $\widetilde{\mathbf{C}}\mathbf{C}^{-1} = \widetilde{\mathbf{B}}\mathbf{B}^{-1}$.

4. Reformulation for linear systems of equations: OrthoRes-type recurrences. Let us now turn from the fixed point equation paradigm to the paradigm for $\Psi\mathbf{x} = \mathbf{o}$, which is still a possibly nonlinear equation in a Banach space. Since $\Phi = \mathbf{I} - \Psi$, it is easy to rephrase all we have done sofar in terms of Ψ. In particular, the basic recursion (2.1) of a (k,ℓ)–step method turns into

$$\mathbf{x}_m := \left[\Psi \left(\sum_{i=1}^{\min\{\ell,m\}} \mathbf{x}_{m-i}\gamma_{m-i,m-1} \right) - \sum_{j=1}^{\min\{k^\circ,m\}} \mathbf{x}_{m-j}\beta^\circ_{m-j,m} \right] \frac{1}{\beta^\circ_{m,m}}\,, \qquad (4.1)$$
$$m = 1, 2, \ldots, M(\leq \infty),$$

or

$$\mathbf{x}_m := \left[\Psi\left(\mathbf{y}_{m-1}\right) - \sum_{j=1}^{\min\{k^\circ,m\}} \mathbf{x}_{m-j}\beta^\circ_{m-j,m} \right] \frac{1}{\beta^\circ_{m,m}}\,, \qquad m = 1, 2, \ldots, M, \qquad (4.2)$$

where, in general, $k^\circ :\equiv \max\{k,\ell\}$ and

$$\beta^\circ_{j,m} := \gamma_{j,m-1} - \beta_{j,m} \quad (0 \leq j < m), \qquad \beta^\circ_{m,m} := -\beta_{m,m} \quad (m > 0)\,. \qquad (4.3)$$

Due to cancellation in $\gamma_{j,m-1} - \beta_{j,m}$ it can happen that $k^\circ < \max\{k,\ell\}$.

In matrix notation, $\mathbf{B} :\equiv (\beta_{j,m})_{j,m=0}^{M}$ is replaced by $\mathbf{B}^\circ :\equiv (\beta^\circ_{j,m})_{j,m=0}^{M}$, which is related by[3]

$$\mathbf{B}^\circ = \mathbf{C}\mathbf{S} - \mathbf{B} + \mathbf{e}_0\mathbf{e}_0^{\mathsf{T}}\,. \qquad (4.4)$$

$\mathbf{B}^\circ$ is a nonsingular upper triangular matrix with column sums 0 (indicated by $^\circ$), which has upper bandwidth k° if this number is finite. So, in particular, $\mathbf{e}^{\mathsf{T}}\mathbf{B}^\circ = \mathbf{o}^{\mathsf{T}}$. Hence, for a (k,ℓ)–step method (2.1) this reformulation has the effect that it becomes a (k°,ℓ)–step method of type (4.1), where $k^\circ \leq \max\{k,\ell\}$. Conversely a (k°,ℓ)–step method of type (4.1) can be turned into a (k,ℓ)–step method (2.1) with $k \leq \max\{k^\circ,\ell\}$. However, generically, equality holds in both inequalities, and therefore, the memory requirements are the same in both paradigms.

The vector

$$\mathbf{r}_m :\equiv -\,\Psi\,\mathbf{x}_m \qquad (4.5)$$

is called the mth residual (associated with the iterate $\mathbf{x}_m$). So, in the recurrence (4.1)–(4.2) we actually compute a residual associated with the auxiliary iterate $\mathbf{y}_{m-1}$.

But when we now turn to the affine case, where

$$\mathbf{r}_m = -\,\Psi\,\mathbf{x}_m = \Phi\,\mathbf{x}_m - \mathbf{x}_m = \mathbf{F}\,\mathbf{x}_m - \mathbf{x}_m + \mathbf{b} = \mathbf{b} - \mathbf{A}\mathbf{x}_m\,, \qquad (4.6)$$

the Ψ-value in (4.1) can be written as a weighted sum of residuals:

$$\mathbf{x}_m := -\left[\sum_{i=1}^{\min\{\ell,m\}} \mathbf{r}_{m-i}\gamma_{m-i,m-1} + \sum_{j=1}^{\min\{k^\circ,m\}} \mathbf{x}_{m-j}\beta^\circ_{m-j,m} \right] \frac{1}{\beta^\circ_{m,m}}\,, \qquad (4.7)$$
$$m = 1, 2, \ldots, M.$$

[3] In [10], $\mathbf{B}^\circ$ was not only chosen as the transposed of the matrix introduced here, but also with opposite sign. The reason is that the residuals were defined with the opposite sign, in contrast to the current standard in the literature on Krylov subspace methods. Moreover, the rank-one correction $\mathbf{e}_0\mathbf{e}_0^{\mathsf{T}}$ that yields here $\beta^\circ_{0,0} :\equiv 0$ was not incorporated, since this element did not appear in the sequel, while here it will appear later in this section.

As usual, due to $\sum_{j=0}^{m-1} \mathbf{b}\,\beta^{\circ}_{j,m} = \mathbf{b}\sum_{j=0}^{m-1}\beta^{\circ}_{j,m} = \mathbf{o}$, it is easy to verify that there is an associated recurrence for the residuals,

$$\mathbf{r}_m := \left[\sum_{i=1}^{\min\{\ell,m\}} \mathbf{A}\mathbf{r}_{m-i}\gamma_{m-i,m-1} - \sum_{j=1}^{\min\{k^{\circ},m\}} \mathbf{r}_{m-j}\beta^{\circ}_{m-j,m}\right]\frac{1}{\beta^{\circ}_{m,m}}\,, \qquad m = 1, 2, \ldots, M. \tag{4.8}$$

In the next step we introduce a "formal matrix" for the residuals,

$$\mathbf{R}_{M+1} :\equiv \left[\begin{array}{cccc} \mathbf{r}_0 & \mathbf{r}_1 & \ldots & \mathbf{r}_M \end{array}\right],$$

so that we can write the recurrence (4.8) for $\mathbf{r}_m$ in "shorthand notation". We have to take into account that in the first sum the column index of $\mathbf{C} = (\gamma_{i,m})_{i,m=0}^{M}$ is shifted by one, a fact that was in (3.6) concealed by the somewhat unusual way of defining $\mathbf{\Phi}(\mathbf{Y})$, while here it requires an extra factor $\mathbf{S}$:

$$\mathbf{A}\mathbf{R}_{M+1}\mathbf{C}\mathbf{S} = \mathbf{R}_{M+1}\mathbf{B}^{\circ}\,. \tag{4.9}$$

This looks similar to a Hessenberg relation $\mathbf{A}\mathbf{R}_M = \mathbf{R}_{M+1}\underline{\mathbf{H}}_M$ with an extended Hessenberg matrix $\underline{\mathbf{H}}_M$, but we still need some adjustments. We note that

$$\mathbf{C}\mathbf{S} = \left[\begin{array}{c|c} \mathbf{o} & \mathbf{C}_M \\ \hline 0 & \mathbf{o}^{\mathsf{T}} \end{array}\right],$$

where $\mathbf{C}_M$ is the $M \times M$ leading principal submatrix of the $(M+1)\times(M+1)$ upper triangular matrix $\mathbf{C}$. Hence, $\mathbf{C}_M$ is also an ST-matrix, which is in accordance with the fact that in (4.4) $\mathbf{B}^{\circ}$ has column sums 0. On the right-hand side of (4.9) the first column of $\mathbf{B}^{\circ}$ is also zero (due to the rank-one modification in (4.4)), so that we can delete the first column on both sides. Writing

$$\mathbf{B}^{\circ} = \left[\begin{array}{c|c} \mathbf{o} & \underline{\mathbf{H}}^{\circ}_M \end{array}\right] \tag{4.10}$$

with an irreducible, $(M+1)\times M$ extended Hessenberg matrix $\underline{\mathbf{H}}^{\circ}_M$, we get for the residual recurrence the alternative "shorthand notation"

$$\mathbf{A}\mathbf{R}_M\mathbf{C}_M = \mathbf{R}_{M+1}\underline{\mathbf{H}}^{\circ}_M\,, \tag{4.11}$$

which finally can be transformed into

$$\mathbf{A}\mathbf{R}_M = \mathbf{R}_{M+1}\underline{\mathbf{H}}^{\blacktriangledown}_M\,, \qquad \text{where} \quad \underline{\mathbf{H}}^{\blacktriangledown}_M :\equiv \underline{\mathbf{H}}^{\circ}_M\mathbf{C}_M^{-1} \tag{4.12}$$

and where $^{\blacktriangledown}$ is supposed to indicate that $\underline{\mathbf{H}}^{\blacktriangledown}_M$ is typically no longer banded when k and ℓ are finite. This is now a Hessenberg relation. But we need to stress that in contrast to (4.11) it is no longer the direct "shorthand notation" of our original residual recurrence, but a consequence of it. In fact, it describes the residual recurrence of an equivalent $(\infty,1)$–step method, which may be, but is unlikely to be a $(k,1)$–step method for some finite k. Neither $\underline{\mathbf{H}}^{\blacktriangledown}_M$ nor the inverse $\mathbf{C}_M^{-1}$ is used in or needed for the original recurrence (4.1). The generalized Hessenberg relation (4.11) with the

extended Hessenberg pencil $(\underline{\mathbf{H}}_M^\circ, \mathbf{C}_M)$ is what represents the original residual recurrence, and it is this relation that we have been looking for. It may describe short recurrences when those of (4.12) are long. This is the main point of (k,ℓ)–step methods. In fact, $\mathbf{C}_M$ has upper bandwidth $\ell-1$ like $\mathbf{C}$, and $\underline{\mathbf{H}}_M^\circ$ has upper bandwidth $k-1$ since the diagonal of $\mathbf{B}_M$ has turned into the lower bidiagonal of $\underline{\mathbf{H}}_M^\circ$.

Note that $\mathbf{e}^\mathsf{T}\mathbf{B}^\circ = \mathbf{o}^\mathsf{T}$ implies that $\mathbf{e}^\mathsf{T}\underline{\mathbf{H}}_M^\circ = \mathbf{o}^\mathsf{T}$ and $\mathbf{e}^\mathsf{T}\underline{\mathbf{H}}_M^\blacktriangledown = \mathbf{e}^\mathsf{T}\underline{\mathbf{H}}_M^\circ\mathbf{C}_M^{-1} = \mathbf{o}^\mathsf{T}$. Hessenberg matrices with this property are said to be of OrthoRes-type, since they define a residual recurrence that is typical for an "idealized generalized conjugate gradient acceleration procedure" of type OrthoRes, as defined in [35].

Conversely: given an irreducible extended Hessenberg matrix $\underline{\mathbf{H}}_M^\blacktriangledown$ with column sums 0, any $M\times M$ ST-matrix $\mathbf{C}_M$ yields by setting $\underline{\mathbf{H}}_M^\circ := \underline{\mathbf{H}}_M^\blacktriangledown\mathbf{C}_M$ an (∞,∞)–step method, which may or may not be a (k,ℓ)–step method for some finite k and ℓ.

Due to the zero column sum property the "shorthand notation" for the recurrences (4.7) for the iterates $\mathbf{x}_n$ are immediately obtained from the one for the residuals by inserting the identity $\mathbf{R}_{M+1} = \mathbf{b}\mathbf{e}^\mathsf{T} - \mathbf{A}\mathbf{X}_{M+1}$ into (4.11), subtracting $\mathbf{b}\mathbf{e}^\mathsf{T}\underline{\mathbf{H}}_M^\circ = \mathbf{o}^\mathsf{T}$, and multiplying from the left by $-\mathbf{A}^{-1}$ to obtain

$$\mathbf{X}_{M+1}\underline{\mathbf{H}}_M^\circ = -\mathbf{R}_M\mathbf{C}_M\,. \tag{4.13}$$

Further note that the two extreme cases in the fixed point paradigm, namely the $(\infty,1)$–step methods with $\mathbf{B}=\widehat{\mathbf{B}}$ and $\mathbf{C}=\mathbf{I}$ and the $(1,\infty)$–step methods with $\mathbf{B}=\mathbf{I}$ and $\mathbf{C}=\widehat{\mathbf{C}}$ now translate into methods with

$$\mathbf{B}^\circ = \mathbf{S} - \widehat{\mathbf{B}} + \mathbf{e}_0\mathbf{e}_0^\mathsf{T}\,, \qquad \mathbf{C}_M = \mathbf{I}$$

and

$$\mathbf{B}^\circ = \widehat{\mathbf{C}}\mathbf{S} - \mathbf{I} + \mathbf{e}_0\mathbf{e}_0^\mathsf{T}\,, \qquad \mathbf{C}_M = \widehat{\mathbf{C}}_M\,,$$

respectively. In view of (4.10), these formulas also yield the corresponding extended Hessenberg pencils $(\underline{\mathbf{H}}_M^\circ, \mathbf{C}_M)$.

In [10], a class of stationary (k,ℓ)–step methods that can be brought into the form of (4.11) and (4.13) is constructed in a way where the recurrence coefficient, *i.e.,* the coefficients of the extended pencil $(\underline{\mathbf{H}}_M^\circ, \mathbf{C}_M)$, are drawn from a rational approximation of a conformal mapping of a simply-connected domain in the complex plain that is known to contain no eigenvalues of $\mathbf{A}$. This construction was following the spirit of its time; see [5, Sections 95–97] and, *e.g.,* [4, 24, 25, 26, 27, 29]. For most users who need to solve a large linear system such a process is too complicated, however. Parsons [29] introduced stationary (k,k)-step methods for linear systems of equations. Other subclasses of linear parameter-dependent (k,ℓ)–step methods were discussed in [11].

Of much more interest are methods where $(\mathbf{H}_M^\circ, \mathbf{C}_M)$ is partly or fully determined by a suitable Galerkin or Petrov–Galerkin condition. An example is the promising IDR(s) method [32], which can be seen to be a $(s+1,s+1)$-step method [14]. There the elements of $(\underline{\mathbf{H}}_M^\circ, \mathbf{C}_M)$ are partly determined by a Petrov–Galerkin condition and partly by smoothing factors that come from a local residual minimization condition. In this case, the representation (4.11) justifies the approximation of some of the eigenvalues of $\mathbf{A}$ by the spectrum of the pencil $(\mathbf{H}_M^\circ, \mathbf{C}_M)$, where $\mathbf{H}_M^\circ$ denotes the upper part of $\underline{\mathbf{H}}_M^\circ$. In IDR($s$) it is possible to separate those eigenvalues of the pencil that are determined by the Petrov–Galerkin condition from those imposed by the smoothing factors. To achieve this, the pencil has to be "purified", "eliminated", and "deflated" in order to isolate the relevant eigenvalues that approximate eigenvalues of $\mathbf{A}$; see [18].

It has been shown by Sonneveld and van Gijzen [32], see also [14], that when $s = 1$, IDR(s) is mathematically equivalent to BiCGStab [33]. Consequently, BiCGStab is mathematically equivalent to a $(2, 2)$-step method. However, its standard implementation is not based on the OrthoRes-type recurrences (4.7)–(4.8).

We will return in the next section to the idea of determining all coefficients by a Galerkin condition imposed by residual minimization. This has been investigated by Barth and Manteuffel [3, 2], but it will turn out that the class of short recurrence methods they considered is slightly different from ours.

5. OrthoMin-type and OrthoDir-type recurrences. An OrthoRes recurrence involving only the residuals $\mathbf{r}_n$ and complemented by an analoguous recurrence for the iterates $\mathbf{x}_n$ is just one way of implementing the classical conjugate gradient and conjugate residual methods and their generalizations. Alternatives are OrthoMin recurrences that involve both residual and direction vectors, and OrthoDir recurrences involving only direction vectors [35]. OrthoMin recurrences are the standard for implementing the classical conjugate gradient (CG) method [20] and the classical biconjugate gradient (BiCG) method [23, 6] and are known to yield ultimately a more accurate numerical solution; see [9, 17]. There are well-known relations between the "shorthand notation" for the three types of recurrences; see [1, 12].

We now assume that all the leading principal submatrices $\mathbf{H}^\circ_m$ of the extended Hessenberg matrix $\underline{\mathbf{H}}^\circ_M$ are nonsingular so that $\underline{\mathbf{H}}^\circ_M$ has an LU decomposition without pivoting,

$$\underline{\mathbf{H}}^\circ_M = \underline{\mathbf{L}}^\circ_M \mathbf{U}_M \,. \tag{5.1}$$

Here, $\underline{\mathbf{L}}^\circ_M$ is lower bidiagonal and has vanishing column sums, that is, it has the form

$$\underline{\mathbf{L}}^\circ_M \equiv: \begin{bmatrix} \omega_0^{-1} & & & & & \\ -\omega_0^{-1} & \omega_1^{-1} & & & & \\ & -\omega_1^{-1} & \omega_2^{-1} & & & \\ & & \ddots & \ddots & & \\ & & & -\omega_{M-2}^{-1} & \omega_{M-1}^{-1} & \\ & & & & -\omega_{M-1}^{-1} & \end{bmatrix} . \tag{5.2}$$

$\mathbf{U}_M$ is unit upper triangular and of upper bandwidth $k-1$. Let us denote its elements by $\psi_{k,n}$:

$$\mathbf{U}_M \equiv: \begin{bmatrix} 1 & \psi_{0,1} & \cdots & \psi_{0,k-1} & 0 & \cdots \\ & 1 & \ddots & \vdots & \ddots & \ddots \\ & & \ddots & \psi_{k-2,k-1} & & \psi_{M-k,M-1} \\ & & & 1 & \ddots & \vdots \\ & & & & \ddots & \psi_{M-2,M-1} \\ & & & & & 1 \end{bmatrix} . \tag{5.3}$$

We now define search directions $\mathbf{v}_n$ (or, direction vectors) and a corresponding matrix $\mathbf{V}_M$ by

$$\begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \cdots & \mathbf{v}_{M-1} \end{bmatrix} :\equiv \mathbf{V}_M :\equiv \mathbf{R}_M \mathbf{C}_M \mathbf{U}_M^{-1} \,, \tag{5.4}$$

so that (4.11), $\mathbf{A}\mathbf{R}_M\mathbf{C}_M = \mathbf{R}_{M+1}\underline{\mathbf{H}}^\circ_M$, becomes equivalent to the coupled identities

$$\mathbf{V}_M\mathbf{U}_M = \mathbf{R}_M\mathbf{C}_M \,, \qquad \mathbf{A}\mathbf{V}_M = \mathbf{R}_{M+1}\underline{\mathbf{L}}^\circ_M \,. \tag{5.5}$$

The second one implies further that the iterates satisfy

$$\mathbf{V}_M = -\mathbf{X}_{M+1}\underline{\mathbf{L}}^{\circ}_M\,, \tag{5.6}$$

which explains the terminology "direction vectors" or "search directions", see (5.7c) below. Eqs. (5.5) and (5.6) are the "shorthand notation" for the OrthoMin-type recursions of a linear (k,ℓ)–step method.

If the assumed factorization (5.1) of $\underline{\mathbf{H}}^{\circ}_M$ does not exist, there is no OrthoMin-type realization of the (k,ℓ)–step method considered. This fact is analogous to a pivot breakdown of the BiCG method, see [13].

Written out in full, these OrthoMin-type recursions are

$$\mathbf{r}_m := \mathbf{r}_{m-1} - \mathbf{A}\mathbf{v}_{m-1}\omega_{m-1}\,, \tag{5.7a}$$

$$\mathbf{v}_m := \sum_{i=0}^{\min\{\ell-1,m\}} \mathbf{r}_{m-i}\gamma_{m-i,m} - \sum_{j=1}^{\min\{k-1,m-1\}} \mathbf{v}_{m-j}\psi_{m-j,m}\,, \tag{5.7b}$$

$$\mathbf{x}_{m+1} := \mathbf{x}_m + \mathbf{v}_m\omega_m\,, \tag{5.7c}$$

$$m = 1,\ldots,M-1.$$

At the start we use $\mathbf{v}_0 := \mathbf{r}_0$ and $\mathbf{x}_1 := \mathbf{x}_0 + \mathbf{v}_0\omega_0$.

Finally, let us turn to OrthoDir-type recursions: eliminating $\mathbf{R}_{M+1}$ from (5.5) and defining

$$\underline{\mathbf{H}}^{\blacktriangle}_M :\equiv \mathbf{U}_{M+1}\mathbf{C}^{-1}_{M+1}\underline{\mathbf{L}}^{\circ}_M \tag{5.8}$$

yields

$$\mathbf{A}\mathbf{V}_M = \mathbf{V}_{M+1}\underline{\mathbf{H}}^{\blacktriangle}_M\,, \tag{5.9}$$

which, together with (5.6), defines the $(\infty,1)$–step OrthoDir-type recursions of a linear (k,ℓ)–step method. However, if $\mathbf{C}_{M+1}$ is not diagonal, *i.e.*, if $\ell > 1$, the matrix $\underline{\mathbf{H}}^{\blacktriangle}_M$ cannot be expected to be banded, since the inverse of the banded triangular matrix $\mathbf{C}_{M+1}$ is, in general, not banded. Worse, unlike in the OrthoRes case, the factor $\mathbf{C}^{-1}_{M+1}$ in (5.8) cannot be brought on the other side of (5.9) since $\mathbf{C}^{-1}_{M+1}$ and $\underline{\mathbf{L}}^{\circ}_M$ do not commute. So, in general, there is no short OrthoDir-type recursion of a (k,ℓ)–step method.

In contrast, if $\ell = 1$ and thus $\mathbf{C}_M = \mathbf{I}$, which is the case for most well-known Krylov space methods, $\underline{\mathbf{H}}^{\blacktriangle}_M$ has the same bandwidth as $\underline{\mathbf{H}}^{\blacktriangledown}_M = \underline{\mathbf{H}}^{\circ}_M$, that is, the upper bandwidth $k-1$. In other words, the basic recurrence of the OrthoDir implementation has the same memory requirement as the basic recurrence of the OrthoRes implementation. However, the need for $\mathbf{r}_m$ for judging the convergence of the method requires an additional recurrence for the residuals and storage for those that are needed, or, alternatively, to compute $\mathbf{r}_m$ according to its definition by spending an extra matrix-vector product in each step.

But let us now *assume* that $\underline{\mathbf{H}}^{\blacktriangle}_M$ can be factored into

$$\underline{\mathbf{H}}^{\blacktriangle}_M \equiv: \underline{\mathbf{H}}'_M(\mathbf{C}'_M)^{-1}\,, \tag{5.10}$$

where $\underline{\mathbf{H}}'_M$ is an irreducible extended Hessenberg matrix of upper bandwidth $k'-1 \geq 0$ and $\mathbf{C}'_M$ is an upper triangular matrix of upper bandwidth $\ell'-1 \geq 0$. Then we obtain OrthoDir-type (k',ℓ')–step recursions described by

$$\mathbf{A}\mathbf{V}_M\mathbf{C}'_M = \mathbf{V}_{M+1}\underline{\mathbf{H}}'_M \tag{5.11}$$

and (5.6). Written out in full, with $\mathbf{C}'_M = (\gamma'_{k,n})$ and $\underline{\mathbf{H}}'_M = (\eta'_{k,n})$, (5.11) becomes

$$\mathbf{v}_{m+1} := \left[\sum_{i=0}^{\min\{\ell'-1,m\}} \mathbf{A}\mathbf{v}_{m-i}\gamma'_{m-i,m} - \sum_{j=0}^{\min\{k'-1,m\}} \mathbf{v}_{m-j}\eta'_{m-j,m} \right] \frac{1}{\eta'_{m+1,m}}, \quad (5.12)$$
$$m = 0, 1, \ldots, M-1.$$

If we let $(s,t) :\equiv (k'-1, \ell'-1)$, then (5.12) is what Barth and Manteuffel [3, 2] call a single (s,t)–recursion. They pointed out that the pair of coupled recursions introduced by Gragg [7] for unitary matrices can be reformulated as a single (s,t)–recursion with $(s,t) = (0,1)$, and that likewise, according to Jagels and Reichel [21, 22] such a recursion with $(s,t) = (1,1)$ is suitable for shifted unitary matrices.

But Barth and Manteuffel were mainly interested in discovering new classes of matrices for which there exists a generalized conjugate gradient methods with error minimization (with respect to a suitable inner product) that can be implemented by a single (s,t)–recursion with $t \geq 1$, that is, $\ell' \geq 2$. The above discussions indicates that the existence of such an optimal (s,t)–recursion is not equivalent to the existence of an optimal (k,ℓ)–step recursion of OrthoRes-type or OrthoDir-type.

6. Conclusions. We have rewritten the (k,ℓ)–step methods of [10] that were originally formulated for systems of fixed point equations $\mathbf{x} = \Phi\mathbf{x}$ as recursions for solving (possibly nonlinear) systems of of equations of the form $\Psi\mathbf{x} = \mathbf{o}$ and, in particular, linear systems of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$. In the latter case, (k,ℓ)–step methods are particular Krylov space solvers, and, according to their original definition, they are of OrthoRes-type. We have transformed them into OrthoMin-type and OrthoDir-type recursions, but discovered in the latter case that short recurrences do not persist in general. Similar short OrthoDir-type recursions have been introduced and investigated by Barth and Manteuffel [3, 2], but, in general, these are not mathematically equivalent to (k,ℓ)–step methods.

REFERENCES

[1] S. F. Ashby and M. H. Gutknecht, *A matrix analysis of conjugate gradient algorithms*, in Advances in Numerical Methods for Large Sparse Sets of Linear Systems, M. Natori and T. Nodera, eds., no. 9 in Parallel Processing for Scientific Computing, Keio University, Yokahama, Japan, 1993, pp. 32–47.

[2] T. Barth and T. Manteuffel, *Multiple recursion conjugate gradient algorithms part i: Sufficient conditions*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 768–796.

[3] T. L. Barth, *Implementation of the conjugate gradient method using short multiple recursions*, PhD thesis, University of Colorado at Denver, 1996.

[4] M. Eiermann, W. Niethammer, and R. S. Varga, *A study of semiiterative methods for nonsymmetric systems of linear equations*, Numer. Math., 47 (1985), pp. 505–533.

[5] D. K. Faddeev and V. N. Faddeeva, *Vycislitel'nye metody linejnoj algebry*, Fizmatgiz, Moscow, 1960. German translation: Numerische Verfahren der linearen Algebra, Oldenbourg, Munich, 1964; 4th ed., 1976. This is not the same book as *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963.

[6] R. Fletcher, *Conjugate gradient methods for indefinite systems*, in Numerical Analysis, Dundee, 1975, G. A. Watson, ed., vol. 506 of Lecture Notes in Mathematics, Springer, Berlin, 1976, pp. 73–89.

[7] W. B. Gragg, *Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle (in Russian)*, in Numerical Methods in Linear Algebra, E. Nikolaev, ed., Moscow University Press, 1982, pp. 16–32. English version, see [8].

[8] ———, *Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle*, J. Comput. Appl. Math., 46 (1993), pp. 183–198.

[9] A. Greenbaum, *Estimating the attainable accuracy of recursively computed residual methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 535–551.

[10] M. H. Gutknecht, *Stationary and almost stationary iterative (k,l)-step methods for linear and nonlinear systems of equations*, Numer. Math., 56 (1989), pp. 179–213.

[11] ———, *On certain types of (k,l)-step methods for solving linear systems of equations*, in Iterative Methods in Linear Algebra, R. Beauwens and P. de Groen, eds., Proceedings IMACS Symposium, Brussels, 1991, 1992, Elsevier (North-Holland), pp. 373–380.

[12] ———, *Changing the norm in conjugate gradient type algorithms*, SIAM J. Numer. Anal., 30 (1993), pp. 40–56.

[13] ———, *Lanczos-type solvers for nonsymmetric linear systems of equations*, Acta Numerica, 6 (1997), pp. 271–397.

[14] ———, *IDR explained*, ETNA, 36 (2010), pp. 126–148.

[15] M. H. Gutknecht and A. Kaiser, *Iterative k-step methods for computing possibly repulsive fixed points in Banach spaces*, J. Math. Anal. Appl., 125 (1987), pp. 104–123.

[16] M. H. Gutknecht, W. Niethammer, and R. S. Varga, *k-step iterative methods for solving nonlinear systems of equations*, Numer. Math., 48 (1986), pp. 699–712.

[17] M. H. Gutknecht and Z. Strakoš, *Accuracy of two three-term and three two-term recurrences for Krylov space solvers*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 213–229.

[18] M. H. Gutknecht and J.-P. M. Zemke, *Eigenvalue computations based on IDR*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 283–311.

[19] K. Hessenberg, *Behandlung linearer Eigenwertaufgaben mit Hilfe der Hamilton-Cayleyschen Gleichung*, Numerische Verfahren, Bericht 1, Institut für Praktische Mathematik (IPM), Technische Hochschule Darmstadt, July 1940. The scanned report and a biographical sketch of Karl Hessenberg's life are available at `http://www.hessenberg.de/karl1.html`.

[20] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 409–435.

[21] C. Jagels and L. Reichel, *The isometric Arnoldi process and an application to iterative solution of large linear systems*, in Iterative Methods in Linear Algebra, R. Beauwens and P. de Groen, eds., Proceedings IMACS Symposium, Brussels, 1991, 1992, Elsevier (North-Holland), pp. 361–369.

[22] C. F. Jagels and L. Reichel, *A fast minimal residual algorithm for shifted unitary matrices*, Numer. Linear Algebra Appl., 1 (1994), pp. 555–570.

[23] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 33–53.

[24] W. Niethammer, *Iterationsverfahren und allgemeine Euler-Verfahren*, Math. Z., 102 (1967), pp. 288–317.

[25] W. Niethammer, *Konvergenzbeschleunigung bei einstufigen Iterationsverfahren durch Summierungsmethoden*, in Iterationsverfahren, Numerische Mathematik, Approximationstheorie (Tagung Iterationsverfahren Numer. Math., Oberwolfach, 1969), Birkhäuser, Basel, 1970, pp. 235–243. Internat. Schr. Numer. Math., Vol. 15.

[26] W. Niethammer and W. Schempp, *On the construction of iteration methods for linear equations in banach spaces by summation methods*, Aequationes Math., 5 (1970), pp. 273–284.

[27] W. Niethammer and R. S. Varga, *The analysis of k-step iterative methods for linear systems from summability theory*, Numer. Math., 41 (1983), pp. 177–206.

[28] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[29] B. Parsons, *General k-part stationary iterative solutions to linear systems*, SIAM J. Numer. Anal., 24 (1987), pp. 188–198.

[30] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd. ed., 2003.

[31] Y. Saad and M. H. Schultz, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comp., 44 (1985), pp. 417–424.

[32] P. Sonneveld and M. B. van Gijzen, *IDR(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations*, SIAM Journal on Scientific Computing, 31 (2008), pp. 1035–1062.

[33] H. A. van der Vorst, *Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644.

[34] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962. Rev. 2nd ed., Springer-Verlag, 1999.

[35] D. M. Young and K. C. Jea, *Generalized conjugate-gradient acceleration of nonsymmetrizable iterative methods*, Linear Algebra Appl., 34 (1980), pp. 159–194.